2011 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 25-30, 2011. San Francisco, CA, USA

Bi-Manual Robotic Paper Manipulation Based on Real-Time Marker Tracking and Physical Modelling

Christof Elbrechter*, Robert Haschke* and Helge Ritter*

Abstract— **The ability to manipulate deformable objects, such as textiles or paper, is a major prerequisite to bringing the capabilities of articulated robot hands closer to the level of manual intelligence exhibited by humans. We concentrate on the manipulation of paper, which affords us a rich interaction domain and that has not yet been solved for anthropomorphic robot hands. A key ability needed for this is the robust tracking and modelling of paper under conditions of occlusion and strong deformation. We present a marker based framework that realizes these properties robustly and in real-time. We compare a purely mathematical representation of the paper manifold with a soft-body-physics model and demonstrate the use of our visual tracking method to facilitate the coordination of two anthropomorphic 20 DOF Shadow Dexterous Hands while they grasp a flat-lying piece of paper, using a combination of visually guided bulging and pinching.**

I. Introduction

The challenge of grasping and manipulating objects with multi-fingered robot hands has seen much progress in recent years. A major factor enabling the handling of a wide variety of objects has been the extensive use of prior off-line simulations to predict stable grasp configurations and contact geometries on the basis of rigid and a-priori known object shapes.

However, many objects, such as textiles, food and paper, are non-rigid, making their shape difficult or impossible to predict. This makes it extremely hard or impossible to adopt geometry-based planning methods to the handling of such objects with robot hands. Consequently, robotic manipulation of highly deformable objects is largely still an unsolved task, calling for the development of strategies that can replace or augment geometry-based planning approaches.

An interesting, useful and rich domain for studying grasping and manipulation strategies for such situations is the manipulation of paper. We believe that a thorough understanding of manipulation strategies for paper and their replication in multi-fingered robot hands will be a significant step towards a synthesis of the "manual intelligence" [1] that we see at work when handling non-rigid objects with our own, human hands.

We focus here on the control of multi-fingered, anthropomorphic hands in a bi-manual setting. In the absence of reliably predictable geometry information, rich sensory feedback becomes an essential source of information to drive the manipulation process and to replace off-line simulation by a closed perception-action loop. Given the limitations of current haptic sensing abilities on anthropomorphic robot hands, our focus in this paper is on vision and we present a method that enables robust and fast tracking of shape changes to a sheet of paper under conditions of occlusion and strong deformation. To allow complex manipulations of the paper, resulting in possible severe occlusions, our approach utilizes fiducial markers [2] covering both sides of the sheet of paper to simplify the visual tracking task. Future work will be aimed at relaxing this limitation and complementing visual information with tactile information from haptic sensors in the finger tips of the robot's hands. A novel design of fiducial markers is presented that is optimized for the paper tracking task, and two different modelling approaches for transforming the visually extracted position information into a 3D model of the paper are compared.

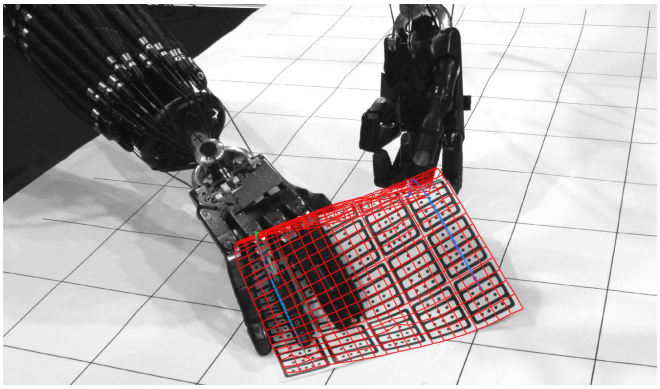


Fig. 1. Two 20 DOF Shadow dexterous robot hands are used to pick up a detected and modelled sheet of paper.

The paper is organized as follows: in Section II we provide an overview of related work. Section III describes the robot and vision setup. Section IV introduces the proposed marker tracking and identification methods. In Section V two different paper modelling techniques are presented and compared. As a non-trivial interaction example, in Section VI we demonstrate a bi-manual strategy for picking up a piece of paper from a flat surface, using visually guided control and combining two interaction primitives, namely "bulging" and "pinch grasp" (see Fig. 1). Finally, Section VII summarizes our results and provides an outlook of future work.

II. Related Work

Robotic paper manipulation has often been addressed from the perspective of origami folding. Balkom and Mason's work on "Robotic Origami Folding" [3] provides a fundamental basis for robotic paper manipulation by formulating difficulty levels for origami models and giving theorems about foldability. They demonstrated their theoretical work

* University of Bielefeld, Neuroinformatics Group, CITEC

on a specialized robot made for folding paper. Mathematical simulation of origami has been a field of research for decades [4]. However, these models are very complex and hard to extend [5]. Our work, in contrast, focuses on the use of generic anthropomorphic hands, albeit initially on a simpler task, which results in a large set of additional challenges. Even picking up a sheet of paper from a flat surface presents us with many unsolved problems.

Simulation of paper and paper-like materials is also a common problem in computer graphics. State of the art approaches use thin shell models based on discrete differential geometry that allow for impressive physical modelling of flexible materials [6], [7]. However, all these modelling techniques lack a link back to the real world to allow tracking of a real sheet of paper, which is required for robotic manipulation. A major challenge in establishing this link is solving the correspondence problem, i.e. mapping image pixels to points on the paper model.

Humans make use of tactile feedback, but currently available tactile sensors are not yet able to give robust and sufficiently sensitive feedback. Anthropomorphic systems that are able to detect, model and manipulate deformable materials have been reported only recently. Kita et al. [8] presented a robot system which was able to spread clothes. Based on dense depth information from a trinocular camera, they fit a physical model to represent the object surface. A similar system that folded towels focused on the visual detection and grasping of towel corners [9]. Even though these approaches produced impressive results, the employed visual detection systems cannot simply be ported to dexterous manipulation due to the presence of occlusions.

Marker aided object detection and tracking is a common technique to circumvent vision problems of occlusions. There are several free marker tracking toolkits available [2], [10]–[12], but none of the existing toolkits were appropriate for our task. The reason being that they do not provide multi-camera pose estimation and are usually not optimized for detecting large numbers of markers within megapixel images in real-time. Most of these libraries also allow for the combination of markers in order to increase pose estimation accuracy, but so far we found no system that uses markers on deformable objects such as cloth or paper. Pilet et al. [13] used feature based key-point registration techniques for augmenting deformable surfaces. We also considered the idea of using SIFT or SURF image features instead of markers, but these approaches are only applicable to rich scenes providing enough structure for robust feature tracking.

III. System Overview

The Bielefeld "Curious Robot" Setup [14] combines visual attention and speech recognition with dexterous bi-manual manipulation skills and proactive dialog communication, which allows robotic manual interaction tasks to be communicated to the robot by lay people. For manipulation tasks, the setup employs two redundant 7-DOF Mitsubishi PA-10 robot arms with two 20 DOF Shadow Dexterous Hands, giving a total of 54 DOF. The robot arms are mounted from the

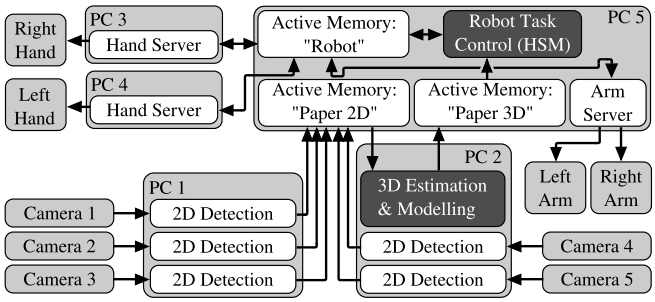


Fig. 2. Data flow and inter process communication (IPC) structure of the setup. IPC is implemented using three event driven "Active Memory" [16] nodes. For each camera a dedicated image capturing and 2D marker detection process feeds the "Paper 2D" node. The "3D Estimation and Modelling" component processes the 2D marker positions into a 3D model of the paper surface, which is committed to the "Paper 3D" node. The "Robot Task Control (HSM)" schedules and monitors the action sequence carried out by the bi-manual robot system to pick up the sheet of paper in a data-driven fashion.

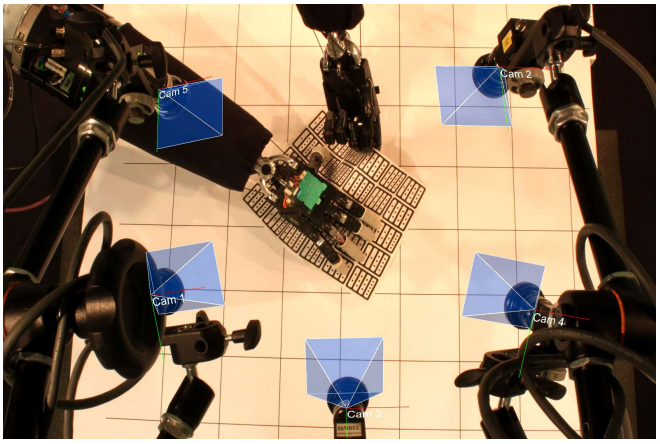
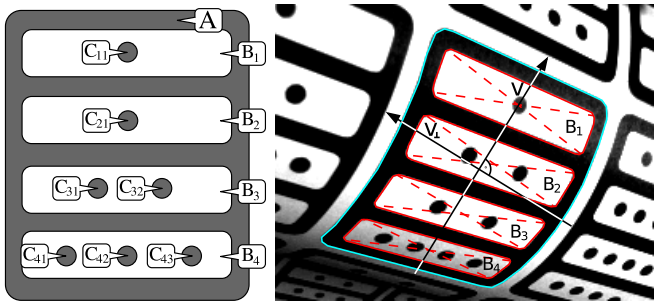


Fig. 3. View from above of the robot interaction area monitored by five calibrated cameras mounted above the areas to avoid collisions with acting hands.

ceiling and controlled by a robot server that handles collision avoidance using an internal scene model.

The Shadow Dexterous Hands [15] are distinguished by their human-like design: in size, number and flexibility of joints, the hands resemble their human counterparts in a thus far unique manner. The Bielefeld hands are equipped with tactile fingertip sensors providing a spatial resolution of 3mm. However, the polyurethane foam, which is required as a pressure-sensitive material, does not provide enough friction for paper manipulation. For this reason, we added rubber finger covers to decrease slippage.

The entire robot system (see Fig. 2) is controlled by several processes distributed over several PCs. The processes communicate and interact using the XCF middleware toolkit [16] which features an event-driven communication scheme. The system's coarse inter-process communication structure is shown in Fig. 2. The described scenario comprises control processes for the robot arms and hands, several vision modules, paper modelling as well as a coordinating hierarchical state machine (HSM) [17]. The vision hardware (see Fig. 3)



(a) (b)

Fig. 4. Fiducial marker design, region layout and identification. **(a)** Example of one of the created fiducial markers. The image shows the top level image region A, 2nd level regions B_{1-4} and 3rd level regions C_{ij}. **(b)** The vector $\mathbf{v}$ is used to sort the B_is. Then $\mathbf{v}_\perp$ is created perpendicular to $\mathbf{v}$ in order to sort the C_{ij}.

consists of five firewire cameras connected to two PCs and running at 15Hz with quad-VGA resolution (1280×960).

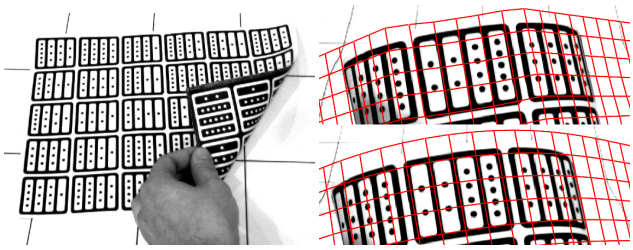
IV. Visual Detection of a Sheet of Paper

Vision based detection of a sheet of paper has become a fundamental component of our system. We use a paper, that is covered on both sides with fiducial markers [2] to facilitate robust and accurate detection. While the usage of less markers, e.g. in the corners of the sheet of paper, might provide adequate results in simple settings, we need full marker coverage to robustly handle strong deformation and occlusion. The grey scale camera images are binarized using an efficient local thresholding operation [18], which increases the robustness of the marker detection in the presence of changing lighting conditions, e.g. due to different alignment towards the light sources. The thresholded images are post-processed using a 3×3 median filter.

The connectivity of image regions is extracted and represented in an containment graph, which can be employed to easily identify the markers [2]. To segment the image into connected regions we employ an approach based on run-length encodings, which was shown to outperform traditional methods by a factor of ten [19]. The source code is available in our open source computer vision library *Image Component Library* (ICL) [20], which was also used for other vision tasks such as camera calibration and 3D visualisation.

A. Fiducial Marker Design and 2D Detection

In order to facilitate the reconstruction of the paper surface, even when a significant number of markers are not detected due to occlusion by hands or the folded paper itself, we need to distinguish a large number of markers distributed on both sides of the paper. To simplify marker discrimination, we created an optimized set of markers that allow for the identification of every sub region of a marker using simple geometrical heuristics. Each marker (see Fig. 4a) consists of a parent region, A, that includes four child-regions, B_i. In turn, each B_i region contains 1 to 5 dot-like sub-regions C_{ij}, where $j \in \{1, N_i\}$ for a given sub region index i. Each marker is uniquely identified by the combination of numbers



(a) (b)

Fig. 5. **(a)** Front and back side 5×6 marker pattern of the used A4 sheet. The marker patterns on the back side is inverted. **(b)** Different kernel bandwidths σ for the plane interpolation model.

N_i of sub-regions C_{ij} at the third level of hierarchy. Thus we denote the marker identity by $M = (N_1, N_2, N_3, N_4)$. For example, the marker in Fig. 4a is defined by $M = (1, 1, 2, 3)$. Hence, M encodes one of $5^4 = 625$ possible unique marker IDs. Since the sheet of paper (and thus the markers) can be arbitrarily rotated w.r.t. the camera, we have to avoid IDs that are ambiguous w.r.t. a reversion of their code M. This was achieved by only allowing IDs with $N_i \leq N_{i+1}$ and $N_1 < N_4$, which reduces the number of possible IDs to 120.

To increase the robustness of the marker detection, we decided to use an error detecting code. In the presence of incorrectly recognized third-level regions C_{ij}, i.e. missing or additional regions, a wrong marker ID is extracted. To allow for the detection of such recognition errors, we further reduced the set of marker IDs to those codes, which pair-wise have a Hamming distance $d(M^1, M^2) = \sum_i |N_i^1 - N_i^2| > 1$. This leaves us a total of 60 remaining markers IDs.

For our experiments, we used an A4 sheet of paper with 5×6 fiducial markers on each side (see Fig. 5a). On the front and back side of the paper we placed identical marker IDs at corresponding positions, which yield identical positions of third-level regions C_{ij} on both sides. However, to distinguish both sides of the paper, the markers on the back side have an inverted color scheme, i.e. using white A and C regions and black B regions (see Fig. 5a). a potential parent region A are processed. If their number of child regions equals four, the number of third-level regions is counted and sorted into a list $M = (N_1, N_2, N_3, N_4)$ in ascending order. If M is a valid marker ID, A is stored in the list of found markers. Please note that this marker detection stage does not exploit geometrical but only topological information.

In the next processing step all marker regions B_i and C_{ij} have to be identified, i.e. ordered geometrically, in order to associate them to corresponding paper model positions in the subsequent 3D processing stage. To this end, a purely topological analysis is not sufficient anymore, e.g. because two adjacent B regions may contain the same number of sub regions and thus cannot be discriminated on this basis alone. Therefore, we employ geometrical information, first estimating the 2D centers of gravity $\mathbf{b}_i$ and $\mathbf{c}_{ij}$ for all B and C regions along with a vector $\mathbf{v}$ (see Fig. 4b). Then, the B_is are sorted according to $\mathbf{b}_i^\top \mathbf{v}$, where $\mathbf{v}$ is the unique vector connecting the two regions $\mathbf{b}_i$ with the largest

Euclidean distance and pointing to the B region with the least number of C regions (see Fig. 4b). The C regions are sorted individually within each B region according to $\mathbf{v}_\perp^\tau \mathbf{c}_{ij}$, where $\mathbf{v}_\perp$ is perpendicular to $\mathbf{v}$. (On the back side of the paper, i.e. when the A region is *white*, $-\mathbf{v}_\perp$ is used. By these means front- and back-side regions are ordered in the same direction, thus associating the same identity to regions lying on opposite sides of the paper.

B. 3D Pose Estimation of the Markers

Based on the correspondences between detected real-world and model-based marker regions, C_{ij}, established in the previous processing stage, we can finally estimate the pose, i.e. the 3D position and orientation, of all markers. To this end, we assume that a single marker can be represented by a planar model, i.e. its bending is negligible. We employ two different approaches to pose estimation, namely a single- and a multi-view approach.

The pose estimation of markers from single camera images is known as *planar pose estimation* and exploits the known marker (model) geometry. Common methods use iterative optimisation methods and require an initial good guess [10], [12], [21]. Yang et al. proposed a closed form solution by reformulating the problem as an homography estimation problem [22]. We also employ this solution because of its simplicity and its low computational complexity.

Obviously, the estimation of depth and out-of-image-plane rotation will be much more robust and accurate, if multiple, non-parallel camera views can be exploited. If a marker is detected in several cameras, we can easily estimate the 3D position of all marker regions C_{ij} using triangulation [23]. Using the estimated world coordinates and corresponding marker-local model coordinates we can easily estimate the marker pose using *absolute orientation* methods [24].

V. Paper Sheet Modelling

In order to model the 3D surface structure of the paper sheet, we compare a purely mathematical approach with a soft-body physics modelling approach. Both models are compared qualitatively based on their modelling accuracy.

A. Soft-Max Interpolation of Detected Marker Planes

In the mathematical approach, the paper surface is modelled as a set of smoothly overlapping plane segments that are associated with the detected markers. Assuming that the markers are small in comparison to the paper curvature, the paper surface in the vicinity of a marker M_i can be approximated by a linear function $\mathbf{F}_i : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ that maps a position $\mathbf{p} \in \mathbb{R}^2$ in paper model coordinates to the corresponding real-world position $\mathbf{w} \in \mathbb{R}^3$. Using the positions of all C regions of a detected marker M_i, we obtain a set of 2D-3D correspondences $\{(\mathbf{p}, \mathbf{w})_{ij} \mid j \in [1 \dots K_i = \sum_k N_{ik}]\}$. As $\mathbf{F}_i$ is assumed to be linear, it can be parametrized by

$$\mathbf{F}_i(\mathbf{p}) = \mathbf{A}_i \, (p_x \;\; p_y \;\; 1)^\tau \, . \tag{1}$$

The parameter matrix $\mathbf{A}_i$ is estimated by standard regression using least-squares fitting, i.e. solving the linear system

$$(\mathbf{w}_{i1} \;\; \mathbf{w}_{i2} \;\; \dots \;\; \mathbf{w}_{iK_i}) = \mathbf{A}_i \begin{pmatrix} p_{i1x} & p_{i2x} & \dots & p_{iK_ix} \\ p_{i1y} & p_{i2y} & \dots & p_{iK_iy} \\ 1 & 1 & \dots & 1 \end{pmatrix} .$$

Finally, the paper surface function $\mathbf{S} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ is created by smooth interpolation between surface patches using soft-max interpolation with Gaussian kernels g_σ of bandwidth σ:

$$\mathbf{S}(\mathbf{p}) = \frac{\sum_i g_\sigma(\mathbf{p} - \mathbf{p}_i)\mathbf{F}_i(\mathbf{p})}{\sum_j g_\sigma(\mathbf{p} - \mathbf{p}_j)} \quad \text{with} \quad g_\sigma(\mathbf{r}) = \mathrm{e}^{-\frac{\|\mathbf{r}\|^2}{2\sigma^2}} \, .$$

B. Physics-based Modelling with the Bullet Physics Engine

Bullet [25] is an open-source physics engine supporting both soft- and rigid body dynamics. We have modelled the sheet of paper as a soft-body object that is defined by a rectangular grid of triangles. The nodes of the physical paper model have to be moved to reflect the visual marker detection results. It turned out that directly setting the position of the nodes or applying forces to them is not feasible due to stability problems. However, we found that setting node velocities based on a simple P-controller – acting on the distance vector from the current to the targeted node position – is a robust method to control the pose of the paper sheet. Every marker (or paper segment) is modelled using 16 quads, each internally split into two triangles. By these means, the mapping between real paper markers and nodes of the physical model becomes simple and there is always one physical node that directly maps to a marker center. We denote $\mathbf{n}^{pos}(\mathbf{p}_i)$ and $\mathbf{n}^{vel}(\mathbf{p}_i)$ as the position and velocity of the physical model node that maps to the marker center at position $\mathbf{p}_i$. After computing the node velocity using the following control law

$$\mathbf{n}^{vel}(\mathbf{p}_i) = K_p(\mathbf{w}_i - \mathbf{n}^{pos}(\mathbf{p}_i)) \, , \tag{2}$$

a physics simulation step is performed. This process is repeated $N = 10$ times during each update cycle, leading to exponential convergence towards the observed marker position $\mathbf{w}_i$. The gain K_p, the time step Δt, and the number N of physics simulation steps per update cycle determine the tracking speed of marker centers. The product $K_p \cdot \Delta t \in (0, 1)$ defines the convergence factor of the control process and we have empirically chosen this factor as 0.9. Choosing smaller values makes the model adaption slower, which reduces the sensitivity to single detection outliers. While the controller directly affects model nodes corresponding to detected marker centers only, the remaining nodes are interpolated smoothly and in a plausible manner by the underlying soft-body physics model.

C. Comparison of both Modelling Techniques

The purely mathematical approach works well as long as a sufficient number of markers are detected (see Fig. 6a-d). The modelling has no temporal memory and constraints, i.e. the paper model is created in a *one-shot* manner. Its single parameter σ can be adapted in order to interpolate between a smooth and a piecewise linear surface (see Fig. 5b).

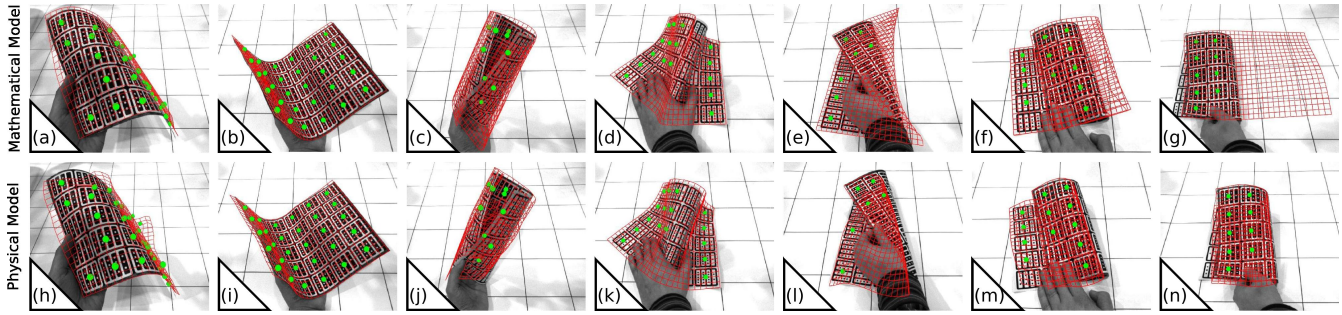


Fig. 6. Comparison of the two modelling techniques. The red grid visualizes the modelled surface. Markers that were actually detected by the vision system are marked with green dots. **(a-g)**: Results of the mathematical model. **(h-n)**: Results of physics based modelling. As long as enough markers are detected and the paper curvature remains small, the mathematical model works well **(a-c)**. Sometimes it is even more convincing than the physical model **(a/h)** and **(c/j)**. However, as soon as a lot of markers are not detected, the lack of distance preservation makes the mathematical model produce unnatural results **(e-g)** whereas the physics based model handles this well **(l-n)**. In **(g)**, the mathematical model does not account for the fact that the sheet of paper was bent against the table top.

However the approach does not take any physical properties of the sheet of paper into account. In particular, the fact that it does not preserve distances on the paper surface makes it behave very unrealisticly in several situations (see Fig. 6e-f). In contrast, the physics based model inherently considers distance preservation, but also collision avoidance, friction properties, and bending constraints. Thus, it behaves comparably in simple situations and better in more complex ones (see Fig. 6k-n). The temporal smoothing introduced by small gain factors K_p of the position controller (2) increases robustness to noisy detection results. Furthermore, the physical model can be more easily extended for modelling further manipulation properties. In particular it is possible to split existing triangles into sub-triangles in order to represent folds [26].

D. Benchmark Results

Single threaded 2D marker detection and identification on grey scale images of size 1280×960 showing all 30 markers takes 14.5ms on an Intel(R) Core(TM)2 Quad Q9550 CPU running at 2.83GHz (local threshold: 7.4ms, median 0.8ms, region detection 2.4ms and region identification and center estimation 3.9ms). This results in a possible theoretical frame rate of 68fps, which is even more than the throughput rate of a firewire bus. In our final setup the framerate drops to 15Hz because up to three cameras share a common IEEE-1394b bus and PC. All 2D detection processes asynchronously feed the 3D modelling process, which runs at a higher frame rate. This allows the 3D model to update, even when a single 2D detection process has provided new data. The mathematical 3D modelling process runs at 50Hz including image acquisition and 3D visualisation. The speed of the physical modelling depends mainly on the number of iterations per update cycle. In our experiments 10 iterations provided good modelling results, resulting in a frame rate of 40Hz.

VI. Bi-manual Paper Picking

Based on the described paper-detection and modelling framework, we implemented a system that allowed our robot to bi-manually pick up a sheet of paper lying on a flat surface. Even though this is a simple every-day task for humans, it requires a complex coordination of hand motion and contact forces as a simple opposing grasp is not feasible. For us, a typical strategy is to lift a corner or an edge by moving our fingernails under the sheet. However, if the paper lies very flat on the table, this can fail. For robotic hands, this strategy is not applicable at all, as sensitive finger nails are missing. Another strategy is to bulge the sheet while closing fingers and thumb towards each other. However, this approach requires a precise control of contact forces within a narrow interval in order to maintain contact and actually move the paper relative to the table surface.

Inspired by the latter approach, we implemented a bi-manual interaction scheme that combines two visually guided motion primitives. Initially, the right robot hand prepares the sheet of paper by bulging it in the middle (see Fig. 7b). Subsequently, the left hand uses a 2-fingered precision grasp to grab the visually identified bulge (see Fig. 7c). To accomplish the first sub task of bulging, the hand is placed above and aligned with the sheet using a central coordinate system attached to the detected paper (see Fig. 7a). After establishing contact with all the finger tips, the bulge is created by a coordinated closing motion of the fingers and an upwards motion of the whole end-effector.

To estimate the topmost position of the bulge, we approximate the outline of the long paper edges by fitting a 4th-order polynomial passing through the marker centers along these edges. The connecting vector between both polynomial's peaks is used to position the left hand. The final grasping phase is the most critical one, because the left hand already lifts the paper during grasping. Consequently, the fixating right hand needs to carefully decrease the contact forces to allow for this lifting motion, while simultaneously maintaining enough force to sustain the bulge. We solved this issue of carefully reducing contact forces using a combination of decreasing the stiffness of hand joints and simultaneously moving the hand upwards.

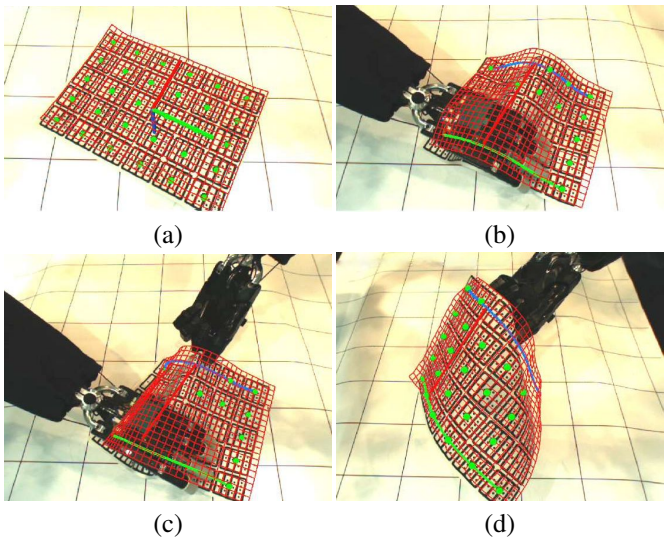


Fig. 7. Example run of the picking up paper experiment: **(a)** The paper sheet is put somewhere into the workspace. Due to the lack of occlusions, all markers are detected (green dots) and the paper model is fitted perfectly. The paper's center coordinate frame is estimated. **(b)** The right robot hand bulges the sheet of paper in order to prepare it for grasping by the left hand. Based on the paper-model, the bulged long edges of the paper are approximated by a simple polynomial model (blue and green arcs). The line between the peaks of these polynomials (thick red line) determines the left hand's grasp point and direction. **(c)** The left hand grasps the front edge of the paper sheet while the right hand still holds it. Then the right hand releases the sheet of paper carefully without dislodging it from the left hand's grasp. **(d)** Finally, the left hand lifts the paper while its shape is still tracked and modelled smoothly.

VII. Conclusions and Future Work

We proposed a strategy for real-time, robust visual detection and physical modelling of a deformable sheet of paper. We demonstrated that the soft-body model of the Bullet physics engine is very suitable for paper modelling. It outperforms the simpler mathematical model not only in tracking accuracy, but is also more promising for future extensions of our technique to paper folding. As a first example of bi-manual paper manipulation using our visual tracking approach, we demonstrated a complex paper-picking task. The implemented system works very robustly independent of different initial paper positions.

Our future work will focus on using the proposed tracking method for more elaborated manipulation actions such as the folding and unfolding of paper. For this, it will be necessary to extend the physical model by explicitly representing folds. Another challenge will be to include tactile feedback in order to robustify the interaction patterns.

The recent advent of cheap and robust 3D cameras, e.g. Kinect, allows us to tackle paper detection and tracking from a new perspective, relaxing the need to cover the sheet of paper with visual markers. However, the correspondence problem, i.e. mapping 3D real-world points onto paper model points, still needs to be solved.

VIII. Acknowledgments

This research was supported by the DFG CoE 277: Cognitive Interaction Technology (CITEC).

References

[1] J. Maycock, D. Dornbusch, C. Elbrechter, R. Haschke, T. Schack, and H. Ritter, "Approaching manual intelligence," *KI - Künstliche Intelligenz*, vol. 24, pp. 287–294, 2010.

[2] M. Kaltenbrunner, "reacTIVision and TUIO: a tangible tabletop toolkit," in *Proc. of the ACM Int. Conference on Interactive Tabletops and Surfaces (ITS).*, 2009.

[3] D. Balkcom and M. Mason, "Robotic origami folding," *The International Journal of Robotics Research*, vol. 27, no. 5, p. 613, 2008.

[4] T. Hull, *Origami3: Third International Meeting of Origami Science, Mathematics, and Education*. AK Peters Ltd, 2002.

[5] T. Tachi, "Simulation of rigid origami," *Origami 4*, p. 175, 2009.

[6] E. Grinspun and A. Secord, "Introduction to discrete differential geometry: the geometry of plane curves," in *ACM SIGGRAPH ASIA 2008 courses*, 2008, pp. 1–4.

[7] E. Grinspun, "A discrete model of thin shells," in *ACM SIGGRAPH 2005 Courses*, 2005, p. 4.

[8] Y. Kita, E. Neo, T. Ueshiba, and N. Kita, "Clothes handling using visual recognition in cooperation with actions," in *Proc. IROS, 2010*, 2010, pp. 2710–2715.

[9] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *Proc. ICRA*, 2010, pp. 2308–2315.

[10] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a Video-Based augmented reality conferencing system," in *Proc. 2nd Int. Workshop on Augmented Reality*, 1999, p. 85.

[11] M. Fiala, "ARTag, a fiducial marker system using digital techniques," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 590–596, 2005.

[12] D. Wagner and D. Schmalstieg, "Artoolkitplus for pose tracking on mobile devices," in *Proc. of 12th Computer Vision Winter Workshop (CVWW'07)*, 2007, pp. 139–146.

[13] J. Pilet, V. Lepetit, and P. Fua, "Augmenting deformable objects in real-time," in *Proc. of the 4th IEEE/ACM Int. Symposium on Mixed and Augmented Reality*, 2005, pp. 134–137.

[14] I. Lütkebohle, J. Peltason, L. Schillingmann, C. Elbrechter, B. Wrede, S. Wachsmuth, and R. Haschke, "The Curious Robot – Structuring Interactive Robot Learning," in *ICRA*, Kobe, 2009.

[15] Shadow Robot Company, "The Shadow Dextrous Hand." [Online]. Available: http://www.shadowrobot.com/hand/overview.shtml

[16] C. Bauckhage, S. Wachsmuth, M. Hanheide, S. Wrede, G. Sagerer, G. Heidemann, and H. Ritter, "The visual active memory perspective on integrated recognition systems," *Image and Vision Computing*, vol. 26, 2008.

[17] D. Harel, "Statecharts: A visual formalism for complex systems," *Science of Computer Programming*, no. 8, pp. 231–274, 1987.

[18] F. Shafait, D. Keysers, and T. Breuel, "Efficient implementation of local adaptive thresholding techniques using integral images," in *Proc. DRR*, vol. 6815. SPIE, 2008.

[19] L. He, Y. Chao, and K. Suzuki, "A run-based two-scan labeling algorithm," *Image Processing, IEEE Transactions on*, vol. 17, no. 5, pp. 749–756, 2008.

[20] C. Elbrechter, M. Götting, and R. Haschke, "Image component library (icl)," http://iclcv.org, 2011.

[21] G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE transactions on pattern analysis and machine intelligence*, p. 20242030, 2006.

[22] Y. Yang, Q. Cao, C. Lo, and Z. Zhang, "Pose estimation based on four coplanar point correspondences," in *Proc. of the 6th int. conference on Fuzzy systems and knowledge discovery - Volume 5*. Tianjin, China: IEEE Press, 2009, pp. 410–414.

[23] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE transactions on robotics and automation*, vol. 12, no. 5, pp. 651–670, 1996.

[24] B. Horn, H. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal Optical Society of America A*, vol. 5, no. 7, pp. 1127–1135, 1988.

[25] "Bullet Physics Library." [Online]. Available: http://www.bulletphysics.org

[26] R. Burgoon, E. Grinspun, and Z. Wood, "Discrete shells origami," *Master's thesis, California Polytechnic State University San Luis Obispo*, 2005.